morphologically analyzing text data;

based on the results of said morphological analysis, generating clauses of said text data;

applying a category dictionary to said clauses to assign concepts (a replacement expression having a representative meaning of the key word) and a category to a key word therein;

generating a syntactic tree of a sentence comprising said clauses according to syntactic tree generation rules;

regarding the key words in said clauses to which a category was assigned, extracting mutually dependent relationships of the key words in the same sentence; and

based on said mutually dependent relationships among the key words, extracting combinations of the categories of the concepts in mutually dependent relationships.

REMARKS

Claims 1-5 have been rejected under 35 U.S.C. §112 as being indefinite because of specifically identified antecedent basis problems. These antecedent basis problems have been corrected.

Claim 1 has been rejected also under 35 U.S.C. §102 as being anticipated by PAIK.

The subject invention pertains to data mining, while PAIK does not pertain to data mining. A key difference is

whether or not a query is needed in order to extract information. In PAIK a query (or browsing request, which is similar) is needed. In the subject invention a query (or browsing request) is not used. It is not necessary for a user of the subject invention to specify query key words or a query category or concept for a search. The information being sought in data mining is information hidden in the data being mined and not document(s) which satisfy a query or browsing request, as in PAIK,

Claim 1 has been amended to make it clear that a query is not needed and to better define "categorized concept" and "unique concept", in order to better distinguish over PAIK, both alone and in combination with other references. PAIK does not teach or suggest extraction of categorized concepts and unique concepts as defined in Claim 1. The claimed concept of extracting a "unique concept" based upon occurrence statistics is not taught or suggested by PAIK.

In accordance with this invention, input text in any form is automatically processed to extract categorized concepts. It is not known by a user, in general, or specified by a user what categorized concepts will be extracted. Then, unique concepts are identified from the extracted categorized concepts by collecting statistical occurrence information for the extracted categorized concepts. In one embodiment, statistical significance is

determined with respect to the extracted categorized concepts in a same category, while in another embodiment two categories are selected and joint occurrence statistics are collected with respect to the extracted categorized concepts in both categories. A GUI display is preferably used to communicate occurrence statistics to a user. Unique concepts are identified as those categorized concepts having a relatively high occurrence rate.

Claim 2 has been rejected under 35 U.S.C. §103 as being unpatentable over PAIK in view of KATAYAMA and KUPIEC. PAIK may be distinguished as described above with respect to Claim 1. KATAYAMA describes a system for identifying a semantic concept and grammatical rule for an undefined word. The relevance of KATAYAMA is not understood. Although KUPIEC has been cited in the rejection as an auxiliary reference, there is no explanation as to how it has been applied to Claim 2. Furthermore, KUPIEC appears to have been filed later than the filing dates of the priority applications and thus would appear not to be valid prior art.

Claims 3 and 4 have been rejected under 35 U.S.C. §103 as being unpatentable over PAIK in view of KUPIEC. PAIK may be distinguished on the same basis as described with respect to Claim 1 and KUPIEC appears to have been filed too late to be applied as valid prior art.

Claims 5 and 6 have been rejected under 35 U.S.C. §103 as being unpatentable over KATAYAMA in view of PAIK. The Examiner argues that KATAYAMA teaches "generating clauses". However, the cited passage (col. 3, lines 25-31) suggests no such thing. "Vocabularies" in this context refers to individual words, not clauses. KATAYAMA furthermore does not teach or suggest use of a category dictionary to assign a category to a key word in a clause as claimed. PAIK also does not teach or suggest extracting combinations of categories. Col 9, line 63 to col. 10, line 4 describes query processing, which extracts documents, not categories. Categories are inputted by a query, not extracted.

## CONCLUSIONS

It is believed that all of the pending claims fully meet all of the requirements of 35 U.S.C. § 112 and also distinguish readily over all of the cited art, when taken individually and in combination. Accordingly, allowance of the pending claims is believed to be in order and is respectfully solicited.

Respectfully submitted,

**JA999-118X**

Ronald L. Drumheller, Attorney
Registration No. 25,674

Ronald L. Drumheller
94 Teakettle Spout Road
Mahopac, NY 10541
Telephone: (845) 628-6090
Facsimile: (845) 628-6197

## VERSION OF CLAIMS SHOWING CHANGES MADE

1. (AMENDED) A data analyzing system for extracting characteristic concepts from [the] data without requiring a query, comprising:

(1) means for extracting categorized concepts from [the data containing] text data, wherein a categorized concept comprises a key word and a category of the key word; and

(2) means for extracting unique concepts from said extracted categorized concepts [extracted from the data], wherein said unique concepts are those extracted categorized concepts which occur conspicuously more frequently within their respective categories than statistically expected [conspicuous within the same content category, the rate occupied by said concepts among the concepts belonging to other corresponding categories exceeding a default].

2. (AMENDED) The system according to claim 1 wherein said means for extracting categorized concepts [(1)] comprises [the] means for:

morphologically analyzing said [the textual part of the] text data;

based on the results of said morphological analysis, generating clauses of said [document] text data;

extracting [any] key words in said clauses as concepts[,] and applying a category dictionary to said clauses to assign [concepts (a replacement expression having a representative meaning of the key word) and] a category to [a] each said key word therein;

analyzing the syntax of a sentence comprising said clauses according to [the] syntactic tree generation rules;

regarding the key words in said clauses to which [concepts and] a category [were] was assigned, extracting mutually dependent relationships of the key words in the same sentence; and

extracting said categorized concepts [, namely] based on said mutually dependent relationships among the key words, and extracting combinations of the categories of the concepts in mutually dependent relationships.

3. (AMENDED) The system according to claim 1 wherein said means for extracting unique concepts [(2)] comprises [the] means for:

receiving an instruction of a user [(means for inputting)];

analyzing said instruction of a user; and

in compliance with said analyzed instruction, presenting said categorized concepts to display with an attribute different from any other concept, of the concepts belonging to the same category, a concept whose statistical characteristic is distinguished beyond a threshold with respect to the set to which it belongs.

4. (AMENDED) The system according to claim 3 wherein said means for extracting unique concepts [(2)] further comprises [the] means for:

calculating the relative frequency of extracted categorized concepts;

searching for categorized concepts from a set of the extracted categorized concepts having a same category;

calculating the frequency of categorized concepts; and

displaying said relative frequency, search results and frequency of concepts that were acquired.

5. (AMENDED) A method for extracting unique concepts from data comprising the phases of:

morphologically analyzing [the textual part of the] text data;

based on the results of said morphological analysis, generating clauses of said [document] text data;

applying a category dictionary to said clauses to assign concepts (a replacement expression having a representative meaning of the key word) and a category to a key word therein;

generating a syntactic tree of a sentence comprising said clauses according to [the] syntactic tree generation rules;

regarding the key words in said clauses to which a category was assigned, extracting mutually dependent relationships of the key words in the same sentence; and

based on said mutually dependent relationships among the key words, extracting combinations of the categories of the concepts in mutually dependent relationships.

6. (AMENDED) A computer-readable record medium recording a program for extracting unique concepts from data, said program including the computer implemented functions of:

morphologically analyzing [the textual part of the] <u>text</u> data;

based on the results of said morphological analysis, generating clauses of said [document] <u>text</u> data;

applying a category dictionary to said clauses to assign concepts (a replacement expression having a representative meaning of the key word) and a category to a key word therein;

generating a syntactic tree of a sentence comprising said clauses according to [the] syntactic tree generation rules;

regarding the key words in said clauses to which a category was assigned, extracting mutually dependent relationships of the key words in the same sentence; and

based on said mutually dependent relationships among the key words, extracting combinations of the categories of the concepts in mutually dependent relationships.